

ЗАДАЧА КЛАСТЕРНОГО АНАЛИЗА БЕЛОРУССКИХ БАНКОВ

А.В. Горбач, студент 5 курса,
кафедра ЭИ и МЭ

Дано множество банков N , $|N|=n$ ($n = 26$), каждый из которых характеризуется по m показателям. Таким образом, каждый банк i характеризуется m -мерным вектором. Каждому показателю k ставится в соответствие бинарное отношение R_k , которое определяет похожест между банками по показателю k . Каждому показателю k ставится в соответствие мера близости ε_k . Тогда бинарное отношение R_k определяется, например, как:

$$R_k = \left\{ (i, j) : |a_{ik} - a_{jk}| \leq \varepsilon_k \right\}.$$

Решением задачи разбиения банков на однородные кластеры будет отношение эквивалентности R , на котором достигается

$$\min \sum_{k=1}^m |R \Delta R_k|,$$

где

$$R \Delta R_k := \{(i, j) : (i, j) \in R, (i, j) \notin R_k \text{ или } (i, j) \notin R, (i, j) \in R_k\}.$$

Используется идея сведения задачи кластерного анализа к задаче дискретной оптимизации следующим образом.

Каждое бинарное отношение R_k представляется с помощью булевой матрицы размерности $n \times n$ следующим образом: для каждого k , $1 \leq k \leq m$, и для каждой пары $(i, j) \in N \times N$ пусть:

$$r_{ij}^{(k)} := \begin{cases} 1, & \text{если } (i, j) \in R_k \\ 0, & \text{если } (i, j) \notin R_k \end{cases},$$

$$r_{ij} := \begin{cases} 1, & \text{если } (i, j) \in R \\ 0, & \text{если } (i, j) \notin R \end{cases}$$

Нелинейная целевая функция $\min \sum_{k=1}^m |R \Delta R_k|$ может быть линеаризована следующим образом:

$$\min \sum_{k=1}^m |R \Delta R_k| = \sum_{k=1}^m \sum_{(i,j) \in N \times N} (r_{ij}^{(k)} - r_{ij})^2 = \sum_{k=1}^m \sum_{(i,j) \in N \times N} r_{ij}^{(k)} + \sum_{k=1}^m \sum_{(i,j) \in N \times N} (1 - 2r_{ij}^{(k)})_{ij} = C + \sum_{(i,j) \in N \times N} C_{ij} r_{ij}$$

$$\text{где } C_{ij} = \sum_{k=1}^m (1 - 2r_{ij}^{(k)}), C = \sum_{(i,j) \in N \times N} (\{k \in \{1..m\} \mid (i, j) \in R_k\}).$$

Искомое бинарное отношение R рефлексивно, симметрично и транзитивно.

Выражая эти требования с помощью линейных уравнений и неравенств, получаем следующую задачу дискретной оптимизации:

$$\min(C + \sum_{(i,j) \in N \times N} C_{ij} r_{ij})$$

$$\begin{cases} r_{ij} = 1, i \in N (\text{рефлексивность}) \\ r_{ij} - r_{ji} = 0, i, j \in N, i \neq j (\text{симметричность}) \\ r_{ij} + r_{jk} - r_{ik} \leq 1, i, j, k \in N (\text{транзитивность}) \\ r_{ij} \in \{0, 1\}, i, j \in N \end{cases}$$

Удалив константу C из целевой функции и исключив часть переменных r_{ij} , в силу $r_{ij}=r_{ji}$, определив новые веса $w_{ij}=c_{ij}+c_{ji}$, получим следующую эквивалентную задачу дискретной оптимизации:

$$\min \sum_{i=1}^n \sum_{j=1}^n w_{ij} x_{ij}$$

$$\begin{cases} x_{ij} + x_{jk} - x_{ik} \leq 1, 1 \leq i < j < k \leq n \\ x_{ij} - x_{jk} + x_{ik} \leq 1, 1 \leq i < j < k \leq n \\ -x_{ij} + x_{jk} + x_{ik} \leq 1, 1 \leq i < j < k \leq n \\ x_{ij} \in \{0, 1\}, 1 \leq i < j \leq n \end{cases}$$

В применяемой далее программе кластерного анализа "Кластер-оптимизация" используется алгоритм секущей плоскости для решения полученной задачи дискретной оптимизации.

Для определения меры близости используем следующую формулу:

$$\varepsilon_k = \frac{\max(a_{ki}) - \min(a_{ki})}{d},$$

где параметр d определяется опытным путем так, чтобы количество кластеров было близко к желаемому.

Эксперимент 1: (кластеризация банков по структуре ресурсов)

Дано 26 белорусских банков, охарактеризованных по 20 показателям: Фонды банка и нераспределенная прибыль; Средства бюджетов; Средства физических лиц; Средства на счетах предприятий и организаций; Внебюджетные средства и фонды, Средства организаций, состоящих на бюджете; Долговые обязательства банка; Корсчета других банков; Депозиты и ресурсы, приобретенные у других банков; Кредиты, полученные от правительства; Кредиты народному хозяйству; Кредиты физическим лицам; Кредиты Правительству и местным органам власти; Ценные бумаги, выпущенные другими юридическими лицами, кроме банков; Собственные долговые ценные бумаги; Средства, перечисленные в фонд обязательных резервов; Денежные средства, золото, драгметаллы; Ресурсы, размещенные в других банках в любой форме; Участие в деятельности других структур (банков, учреждений, организаций); Основные средства и нематериальные активы, т.м.ц.; Корсчета в других банках. Иными словами – это показатели по всем видам валют из отчетности по форме 02056.

По каждому из этих показателей определяем меру близости:

$$\varepsilon_k = \frac{\max(a_{ki}) - \min(a_{ki})}{20}$$

Используя программу "Кластер-оптимизация", получаем решение в виде следующих 8 кластеров:

1. БелПСБ,
2. СБ Беларусбанк,
3. БелВЭБ,
4. Белбизнесбанк,
5. Приорбанк,
6. БелАПБ,
7. МинскКомплексБанк,
8. Поиск, Слафнефтьбанк, Белкомбанк, Мосбизнесбанк, Белгазпром, Белорусский банк развития, Золотой талер, Транзитный банк, Инфобанк, Народный банк, Джембанк, Биржевой банк, Абсолютбанк, Белкоопбанк, Технобанк, РРБ-банк, Новоком, Профбанк, Белбалтия.

Результаты решения позволяют сделать вывод: все системообразующие банки и МинскКомплексБанк, столь непохожие друг на друга, создают отдельные кластеры, а все остальные имеют на их фоне меньшую непохожесть и попадают в один кластер. Для устранения этого недостатка будем сравнивать не значения показателей, а их порядки – в качестве входных данных используем логарифмы от значений показателей, т.е. $a_{ij}^{(k)} := \log(a_{ij}^{(k)} + 1)$.

Эксперимент 2: (кластеризация банков по структуре ресурсов)

Дано: исходные данные такие же как и в эксперименте 1, только делаем замену $a_{ij}^{(k)} := \log(a_{ij}^{(k)} + 1)$.

По каждому из этих показателей определяем меру близости:

$$\varepsilon_k = \frac{\max(a_{ki}) - \min(a_{ki})}{20}$$

Используя программу "Кластер-оптимизация", получаем решение в виде следующих 6 кластеров:

1. БелПСБ, БелВЭБ, Приорбанк.
2. СБ Беларусбанк,
3. Белбизнесбанк, БелАПБ,
4. Поиск, Белкомбанк, Белгазпром, Золотой талер, Транзитный банк, Инфобанк, Биржевой банк, Технобанк,
5. Слафнефтьбанк, Мосбизнесбанк, Белорусский банк развития, Белбалтия
6. Народный банк, Джембанк, Абсолютбанк, Белкоопбанк, РРБ-банк, Новоком, Профбанк.

Эксперимент 3: (кластеризация банков по структуре кредитной политики)

Дано 26 банков, охарактеризованных по 5 показателям (Кредиты народному хозяйству; Кредиты физическим лицам; Кредиты Правительству и местным органам власти; Кредиты другим банкам; Сомнительные, просроченные и пролонгированные кредиты).

По каждому из этих показателей определяем меру близости:

$$\varepsilon_k = \frac{\max(a_{ki}) - \min(a_{ki})}{50}.$$

Используя программу "Кластер-оптимизация", получаем решение в виде следующих 13 кластеров:

1. БелПСБ,
2. СБ БеларусБанк,
3. БелВЭБ,
4. БелБизнесБанк,
5. ПриорБанк,
6. БелАПБ,
7. Поиск,
8. Слафнефтьбанк, Мосбизнесбанк, Золотой талер, Транзитный банк, Инфобанк, Народный банк, Джембанк, Биржевой банк, Абсолютбанк, Белкоопбанк, РРБ-банк, Новоком, Профбанк, Белбалтия,
9. Белкомбанк,
10. Белгазпром,
11. Белорусский банк развития,
12. МинскКомплексБанк,
13. Технобанк.

В действительности не все показатели банка можно считать равнозначными, и поэтому имеет смысл ввести стоимостные коэффициенты c_k . Тогда целевая функция будет иметь вид:

Эксперимент 4: (кластеризация банков по структуре кредитной политики)

Задача эксперимента 3 решается с использованием в качестве исходных данных логарифмов значений показателей.

По каждому из этих показателей определяем меру близости: $\varepsilon_k = \frac{\max(a_{ki}) - \min(a_{ki})}{7}$.

Используя программу "Кластер-оптимизация", получаем решение в виде следующих 7 кластеров:

1. БелПСБ, СБ Беларусбанк, БелВЭБ,
2. Белбизнесбанк, Приорбанк, Белкомбанк, Белорусский банк развития, МинскКомплексБанк,
3. БелАПБ, Поиск, Мосбизнесбанк, Белгазпром, Инфобанк,
4. Слафнефтьбанк, Белбалтия,
5. Золотой талер, Технобанк,
6. Транзитный банк, Народный банк, Биржевой банк, Абсолютбанк, Белкоопбанк, РРБ-банк, Новоком, Профбанк,
7. Джембанк.

Для получения более реальной кластеризации по структуре кредитной политики возможно рассматривать не абсолютные значения показателей, а относительные, например процент от собственного капитала банка.

Эксперимент 5

Кластеризация по следующим показателям:

- Прибыль/Собственный капитал,
- Прибыль/Уставный фонд,
- Прибыль/Баланс,

- Собственный капитал,
- Уставный фонд.

Данные взяты из балансовых отчетов за 1998 год.

Мера близости определена следующим образом: $\varepsilon_k = \frac{\max(a_{ki}) - \min(a_{ki})}{10}$.

Используя программу "Кластер-оптимизация", получаем решение в виде следующих 7 кластеров:

1. СБ Беларусбанк, МинскКомплексБанк, Белорусский банк развития, Поиск,
2. БелВЭБ,
3. БелПСБ, БелАПБ, Приорбанк,
4. Белбизнесбанк,
5. Слафнефтьбанк, Белгазпром, Биржевой банк, Мосбизнесбанк, Инфобанк, Технобанк, Транзитный банк, Золотой талер,
6. Белбалтия, Белкомбанк, РРБ-банк, Абсолютбанк, Белкоопбанк, Новоком, Профбанк, Народный банк,
7. Джембанк.

Эксперимент 6

Кластеризация по следующим показателям:

- Прибыль/Собственный капитал,
- Прибыль/Уставный фонд,
- Прибыль/Баланс.

Данные взяты из балансовых отчетов за 1998 год.

Мера близости определена следующим образом: $\varepsilon_k = \frac{\max(a_{ki}) - \min(a_{ki})}{10}$.

Используя программу "Кластер-оптимизация", получаем решение в виде следующих 7 кластеров:

1. СБ Беларусбанк, БелВЭБ, БелАПБ, МинскКомплексБанк, Белорусский банк развития, Слафнефтьбанк, Белбалтия, Поиск, Белкомбанк, РРБ-банк, Абсолютбанк,
2. БелПСБ, Белгазпром, Биржевой банк,
3. Приорбанк,
4. Белбизнесбанк,
5. Мосбизнесбанк, Инфобанк, Технобанк, Транзитный банк, Золотой талер, Белкоопбанк, Новоком, Народный банк,
6. Джембанк,
7. Профбанк.

В заключение хотелось бы подчеркнуть, что данные результаты носят исключительно иллюстративный характер и требуют более тонкой доводки до получения кластеров в соответствии с международной классификацией.

ЛИТЕРАТУРА

1. М.М. Ковалев, И.Т.Шибeko. Диагностика банка: анализ, рейтинги// Финансовый анализ, 1997, №9-10.
2. M. Grotschel, Y. Wakabayashi. A cutting plane algorithm for a clustering problem.